Some Exact and Approximate Ancillary Statistics

by

Robert J. Buehler *

University of Minnesota

Technical Report No. 438

July 1984

Abstract

A confidence distribution of θ given x can be obtained by inverting the conditional distribution of $\hat{\theta}(x)$ given $u(x)$ where $u(x)$ is either an exact or approximate ancillary. Approximate ancillaries proposed by Fisher, Pierce and Efron and Hinkley are discussed and their relationship to translation and transformation models is discussed. A new ancillary called the predictive ancillary is proposed. It is obtained in closed form for an autoregressive model and is shown by Monte Carlo simulation to contain less information than competitors.

Key words and phrases: Ancillary statistic, approximate ancillary, conditional inference, Fisher information, confidence intervals.

*Research supported by National Science Foundation Grant MCS8202174

1. The problem.

We begin by briefly indicating the point of view which has motivated the present study.

- Confidence intervals are more informative than tests and are therefore preferable.
- Confidence distributions (defined in Section 2) are more informative than confidence intervals and are therefore preferable.
- Confidence procedures suffer from nonuniqueness associated with the choice of reference set. Arbitrariness in the reference set is the counterpart in frequency theory to the Bayesian's arbitrary prior (Lindley, 1971, p. 436).
- Fiducial theory deals with nonuniqueness to a degree, but is limited in scope.
- The correct coverage property of confidence intervals can be achieved by making some intervals too long and others too short. This can happen when using criteria based on power, as shown by Cox (1958).
- Confidence intervals should be neither too long nor too short.
- Optimality criteria for inference should be consistent with a continuity principle which asserts that slight changes in the model should give only slight changes in the inference.
- Completely satisfactory criteria for confidence intervals have yet to be formulated.

In the present paper we review some methods for comparing confidence intervals and confidence distributions. An ancillary statistic or an approximate ancillary can provide a reference set for confidence intervals, and we review some ancillaries which have been proposed. Attention is drawn to the likelihood shape statistic which is exactly ancillary in translation models. A new ancillary is proposed which we call (for want of a better term) the predictive ancillary. It is shown that the predictive ancillary can be calculated in closed form for an autoregressive model and that it is indeed approximately distribution constant as measured by its Fisher information. A further summary of results is given in Section 8.

2. Confidence distributions and induced distributions.

We assume parametric models $f(x;\theta)$, $\theta_U < \theta < \theta_L$. We will say a statistic $u(x)$ is distribution constant if its distribution is the same for all θ. This expression is preferred to ancillary statistic because it is convenient to reserve the latter for more casual usage meaning approximate ancillary, sometimes ancillary, etc. Equivalent to distribution constant would be exactly ancillary or Basu ancillary.

For any γ, $0 < \gamma < 1$, the quantity $\bar{\theta}_\gamma(x)$ is an upper confidence limit for θ with confidence level γ if

$$P_\theta\{\theta < \bar{\theta}_\gamma(x)\} = \gamma \quad \text{for all } \theta. \tag{2.1}$$

If $g(\theta|x)$ is a density of θ for each x and if the γ percentile $\bar{\theta}_\gamma$ defined by

$$\int_{-\infty}^{\bar{\theta}_\gamma(x)} g(\theta|x)\,d\theta = \gamma \tag{2.2}$$

$$g_{t|u}(\theta|x) = -\partial H(t|u;\theta)/\partial\theta \ . \qquad (2.4)$$

This construction gives exact confidence limits whether or not u is exactly ancillary. To see this let $\bar{\theta}_\gamma(x) = \bar{\theta}_\gamma(t,u)$ be the γ percentile of the density $g_{t|u}(\theta|x)$. Then the conditional probability that $\theta < \bar{\theta}_\gamma = \bar{\theta}_\gamma(t,u)$ given u is equal to γ for each u. Therefore the unconditional probability is also γ. This is a relevant fact when dealing with approximate ancillaries. Of course we do need regularity conditions to the effect that $H(t|u;\theta)$ is such that (2.4) is a density function for each x.

If $t_2 = t_2(t_1,u)$ is a monotone function of t_1 for each fixed u, then (t_1,u) and (t_2,u) give the same induced distribution. If u_1 and u_2 are essentially different (not one-to-one) then in general (t,u_1) and (t,u_2) give different induced distributions.

3. Confidence distributions and tests.

If $\bar{\theta}_\gamma(x)$ is an upper confidence limit for θ then for any θ_0 the set $\{x|\theta_0 > \bar{\theta}_\gamma(x)\}$ is a size $\alpha = 1-\gamma$ critical region for testing $H_0:\theta = \theta_0$. Contrariwise if for each θ_0 a size α test of $H_0:\theta = \theta_0$ has critical region $C_\alpha(\theta_0)$ then the set $R_\gamma(x) = \{\theta|x\varepsilon C_\alpha(\theta)\}$ is a confidence region with confidence level $\gamma = 1-\alpha$. If for each fixed x the sets $R_\gamma(x)$ are nested intervals, then the construction yields a confidence distribution.

In the pivotal construction the solution depends on the choice of (t,u). The alternative construction via tests seemingly provides still more possible solutions. Especially in cases where uniformly most powerful tests do not exist, one is faced with arbitrary choices for determining

satisfies (2.1) for each γ, then we have nested confidence intervals and $g(\theta|x)$ will be called a confidence distribution, a term used for example by Box et al (1978), p. 114.

If $t = t(x)$ is any one-dimensional statistic (typically an estimator of θ) with CDF $H(t;\theta)$ then the induced density of θ (based on t) is

$$g_t(\theta|x) = -\partial H(t(x);\theta)/\partial\theta \ . \qquad (2.3)$$

[Regularity conditions are not our main concern, but they are needed here: the partial derivative $\partial H/\partial\theta$ must exist and be negative and $\lim H(t;\theta)$ must equal 0 (or 1) as θ tends to θ_U, its upper limit (or to θ_L, its lower limit).] The distribution corresponding to the induced density is automatically a confidence distribution. The adjective "confidence" suggests the interpretation while "induced" reminds us where it came from.

A pivotal quantity or pivot is a function of x and θ whose distribution is free of θ. Since $H(t;\theta)$ in (2.3) has a uniform (0,1) distribution it is a pivot. The pivotal method of obtaining (2.3) consists in transforming random variable H to random variable θ with t held constant.

If there is a monotone one-to-one relationship between t_1 and t_2 then both give the same induced distribution, but essentially different statistics of course give different induced distributions.

Now let (t,u) be any pair of statistics where t is again one-dimensional but u is unrestricted for the moment. If $H(t|u;\theta)$ is the CDF of t conditional on u then (subject to regularity conditions similar to those just mentioned) the induced density of θ based on t-given-u is defined by

a test. Examples of choices would be: (1) use locally most powerful one-sided test of θ_0 for each θ_0; (2) use most powerful test of θ_0 versus θ_0+1; (3) use most powerful test of θ_0 versus $2\theta_0$; etc.

4. Criteria for comparing induced distributions.

Neyman's approach to optimality of confidence intervals was to link the theory to that of most powerful tests. This leads to concepts such as uniformly most accurate confidence sets (see for example Lehmann (1959), p. 78). Cox's (1958) example of two measuring instruments casts doubt on Neyman's criterion. The problem is that the most powerful test is achieved by making some intervals too long and others too short.

4.1 Comparison by relevant subsets.

Suppose we have two possible solutions $\bar{\theta}_\gamma^{(1)}(x)$ and $\bar{\theta}_\gamma^{(2)}(x)$ obtained from different confidence distributions. For fixed γ put

$C_\gamma = \{x | \bar{\theta}_\gamma^{(1)}(x) < \bar{\theta}_\gamma^{(2)}(x)\}$, C_γ^c = complement of C_γ. A person who favors solution 1 would feel that $\bar{\theta}_\gamma^{(2)}(x)$ is too large in C_γ and too small in C_γ^c. Thus we can check whether C_γ (or C_γ^c) is a positively (or negatively) biased relevant subset (in the sense of Buehler (1959)) for solution 2. Similarly the bias of the same sets can be tested against solution 1. An example is given in Section 6. This approach has the disadvantage of needing two solutions. Given a particular solution it may not be evident what competitor to test it against.

4.2 Comparison by Fisher information.

If our confidence distribution comes from an induced distribution based on statistics (t,u), then it seems reasonable to look at the Fisher information. For this we require not just the usual marginal information but also the conditional information. Suppose the joint density of (t,u) is factored into conditional and marginal according to

$$f(t,u;\theta) = f_1(t|u;\theta)f_2(u;\theta). \tag{4.1}$$

where we deliberately allow a θ in f_2. Let the logarithms of f, f_1, f_2 be ℓ, m_1, m_2, so that if prime denotes derivative with respect to θ,

$$\ell = m_1 + m_2, \quad \ell' = m_1' + m_2', \quad \ell'' = m_1'' + m_2'' . \tag{4.2}$$

Some standard identities that carry over to this situation are

$$\begin{aligned} &E(m_1'|u) = 0, \quad E\ell' = Em_1' = Em_2' = 0 \\ &E(m_1'^2|u) = -E(m_1''|u) \\ &E(\ell'^2) = -E\ell'', \quad E(m_2'^2) = -Em_2'' \quad . \end{aligned} \tag{4.3}$$

Consistent with standard usage we define the information in (t,u) as

$$i_{t,u}(\theta) = -E\ell'' = E\ell'^2, \tag{4.4}$$

the Fisher information in u as

$$i_u(\theta) = -Em_2'' = Em_2'^2, \tag{4.5}$$

and the conditional information in t given u as

$$i_{t|u}(\theta|u) = -E(m_1''|u) = E(m_1'^2|u). \tag{4.6}$$

When u is distribution constant then $m_2' = m_2'' = i_u(\theta) = 0$, and $i_{t|u}(\theta)$ could equivalently have been defined as $-E(\ell''|u)$ as in Cox and Hinkley (1974), p. 110. The Cox-Hinkley definition avoids t, and indeed even in the present case where u need not be distribution constant the t in (4.6) is redundant in the sense that if t_2 is a one-one function of t_1 and u then $i_{t_2|u}(\theta) = i_{t_1|u}(\theta)$.

Now consider the identity $\ell'' = m_1'' + m_2''$. If we take first conditional and then marginal expectation we get

$$i_{tu}(\theta) = Ei_{t|U}(\theta) + i_u(\theta). \tag{4.7}$$

In words, the total information equals the expectation of the conditional information plus the marginal information. When u is distribution constant so that $i_u(\theta) = 0$ this reduces to a well known result of Fisher which purports to explain how the ancillary u recovers the lost information.

Fisher information furnishes some possible criteria for deciding what to condition on and whether to condition at all: (a) u_1 is a better conditioning statistic than u_2 if

$$i_{u_1}(\theta) < i_{u_2}(\theta) \quad \text{for all} \quad \theta. \tag{4.8}$$

(b) Conditional inference using $t|u$ is preferred to unconditional inference using t if

$$Ei_{t|U}(\theta) > i_t(\theta) \quad \text{for all } \theta, \tag{4.9}$$

or equivalently if

$$i_t(\theta) + i_u(\theta) < i_{tu}(\theta) \quad \text{for all } \theta. \tag{4.10}$$

Because of the dependence on θ, these two criteria give only partial orderings.

4.3 Comparison by score function.

Let $A_1,\ldots,A_k$ denote mutually exclusive and exhaustive outcomes and let $p = (p_1,\ldots,P_k)$ be the corresponding probability vector. In a subjective setting a "probability appraiser" is awarded a "score" or "payoff" $g_i(\hat{p})$ if $\hat{p} = (\hat{p}_1,\ldots,\hat{p}_k)$ is his appraised value of p and if A_i occurs (Brier 1950, Good 1952, Savage 1971). The function $g_i(p)$ is "proper" (encourages honesty) if $\Sigma p_i g_i(p) \geq p_i g_i(\hat{p})$. In a continuous setting replace p_i by $f(y)$, $\hat{p}$ by $\hat{f}$, $g_i(p)$ by $g(y,f)$ and $\Sigma p_i g_i(\hat{p})$ by $\int g(y,\hat{f}) f(y)dy$ (Hendrickson and Buehler, 1971). Then $g(y,f) = \log f(y)$, for example, is proper because $\int(\log f(y))\, f(y)\, dy \geq \int(\log \hat{f}(y))\, f(y)dy$ for every pair of densities f, $\hat{f}$. Now let $\hat{f}(\theta|x)$ denote any confidence distribution of θ given x, where θ now plays the role of y. Buehler (1971) suggests measuring the performance of $\hat{f}$ by

$$w(\theta,\hat{f}) = E \log \hat{f}(\theta|x) = \int\{\log \hat{f}(\theta|x)\}\, f(x;\theta)dx \ .$$

A partial ordering of confidence distributions will then result from the criterion: $\hat{f}_1$ is preferred to $\hat{f}_2$ if

$$w(\theta,\hat{f}_1) \geq w(\theta,\hat{f}_2) \quad \text{for all } \theta. \tag{4.11}$$

(Alternatively for $g(x,\hat{f}) = \log \hat{f}(x)$ we could substitute another proper score such as $g(x,\hat{f}) = 2\,\hat{f}(x) - \int\hat{f}^2 dx$.)

5. Some exact and approximate ancillary statistics.

In this section we list a few of the exact and approximate ancillary statistics which have been proposed. After giving the basic definitions we present a tabulation giving basic properties and references.

5.1 The likelihood shape statistic.

The likelihood shape statistic $w(x)$ specifies the shape of the likelihood function apart from its location. More formally, if $f(x;\theta)$ is the likelihood, $\hat{\theta}$ is the MLE, x_1 and x_2 are any two points in the sample space, then we say $w(x)$ is a likelihood shape statistic if

$$w(x_1) = w(x_2) \Leftrightarrow f(x_1;\theta + \hat{\theta}(x_1)) \propto f(x_2;\theta + \hat{\theta}(x_2)). \qquad (5.1)$$

We show in Appendix A that w is exactly ancillary for translation models, but not in general. More casually w can be called the likelihood shape statistic with the usual understanding about an equivalence class of one-to-one functions.

5.2 Fisher's ancillary.

This is defined in the notation of Appendix A as $u_F = \hat{\ell}''$, the second derivative of the log likelihood evaluated at the maximum. Clearly u_F is a component of the likelihood shape.

5.3 Pierce's ancillary.

We define $u_P = -\hat{\ell}''/\hat{i}$ where $\hat{i}$ is the "estimated Fisher information (see (A.7) in Appendix A).

5.4 Efron and Hinkley's ancillary.

We define $u_{EH} = (1 - u_p)/\hat{\gamma} = (\hat{\ell}'' + \hat{i})/\hat{i}\hat{\gamma}$ where $\hat{\gamma}$ is the MLE of Efron's curvature γ (see Appendix A).

5.5 The likelihood ratio ancillary.

If the parameter space $\{\theta\} = \omega$ is imbedded in a larger space Ω then under general conditions it is well known that the likelihood ratio statistic for testing ω versus Ω has a null distribution asymptotically free of θ. This is the basis for using the likelihood ratio statistic as an approximate ancillary, as proposed by Barndorff-Nielsen (1980) in the case of curved exponential families.

5.6 The predictive ancillary.

In Appendix B we suggest an ancillary based on fiducial theory. If X given θ has c.d.f $F_1(x;\theta)$ and Y given $X = x$ and θ has c.d.f. $F_2(y|x;\theta)$ then the predictive ancillary is

$$\Psi(x,y) = \int F_2(y|x;\theta)\,|\partial F(x;\theta)/\partial\theta|\,d\theta \qquad (5.2)$$

An example is given in Section 7.

5.7 The Bayes ancillary.

If $f^{(\pi)}(\theta|x)$ is the posterior density relative to prior π, then the Bayes ancillary (Appendix B) is

$$\Psi^{(\pi)}(x,y) = \int F_2(y|x;\theta) f^{(\pi)}(\theta|x)\,d\theta \ . \qquad (5.3)$$

We hope to give some examples in a later report.

5.8 Properties and references.

Table 1 compares properties of the ancillaries just mentioned and lists some references.

6. A normal example.

Observe x_1, x_2 from $N(\theta,1)$. While trivial in some respects, this example demonstrates a number of things to look for in general. Unconditional inference using the sufficient statistic $\bar{x}$ gives the induced distribution

$$\{\theta|x_1,x_2\} \sim N(\bar{x},1/2). \tag{6.1}$$

Consider conditional inference using (t,u) where $u = ax_1 + bx_2$, $a = \cos\lambda$, $b = \sin\lambda$, $-\pi < \lambda \leq \pi$ (essentially giving us an arbitrary linear function), and where $t = \bar{x}$. As mentioned in Sec. 2, other choices of t give the same induced distribution and the analysis is simpler with $t = bx_1 - ax_2$. The distribution of t given u is then the same as the unconditional distribution,

$$t \text{ or } \{t|u\} \sim N(b-a)\theta,1) \tag{6.2}$$

which gives the induced distribution

$$\theta \sim N\left((b-a)^{-1}(bx_1 - ax_2),\ (b-a)^{-2}\right). \tag{6.3}$$

Here are some things to notice. (1) Central confidence intervals based on (6.3) are wider than those based on (6.1) (because $(b-a)^2 \leq 2$). (2) Upper confidence limits (95 percent, say) using (6.1) could be either larger or smaller than those using (6.3). (3) If $(a,b) = (1,0)$ the

Name of Ancillary	Symbol	Available from Likelihood Function	Exact for Translation Models	Exact for Transformation Models	References
Likelihood shape	w	yes	yes	no	Appendix A
Fisher's	$u_F=I(x)=\hat{\ell}''$	yes	yes	no	Fisher (1934), Efron and Hinkley (1978)
Pierce's	$u_p=-\hat{\ell}''/\hat{i}$	no	yes	yes	Pierce (1975), Efron and Hinkley (1978), Barndorff-Nielsen (1982) Buehler (1982)
Efron and Hinkley's	$u_{EH}=(1-u_p)/\hat{\gamma}$	no	yes	yes	
Likelihood ratio		no	no	no	Barndorff-Nielsen (1980)
Predictive	$\Psi(x,y)$	no	yes	yes	Appendix A
Bayes	$\Psi^{(\pi)}(x,y)$	no	no	no	Appendix A

TABLE 1

solution is the same as we would get if we disregard x_1 and use x_2 only (and vice versa for (0,1)). (4) The degenerate case $a = b = 2^{-1/2}$ throws away everything. (5) If $a = -b = 2^{-1/2}$ the solutions coincide and all information is used. (6) The Fisher information in the marginal distribution of u and in the conditional distribution of t given u equal $(a+b)^2$ and $(a-b)^2$ respectively. These add to 2, the information in (x_1, x_2). This shows the proportion of information lost using (6.3) rather than (6.1). (7) For confidence level γ the upper confidence limits from (6.1) and (6.3) equal respectively

$$\bar{\theta}_{\underline{x}} = \bar{x} + 2^{-1/2} z_\gamma \quad \text{and} \quad \bar{\theta}_{t|u} = (bx_1 - ax_2 + z_\gamma)/(b-a). \quad (6.4)$$

Assuming $b > a$, $\bar{\theta}_{t|u} < \bar{\theta}_{\underline{x}}$ if and only if

$$x_1 - x_2 < \frac{2z_\gamma}{b+a}\,(1 + 2^{-1/2}(b-a)).$$

As mentioned in Section 4.1, a reasonable attitude when comparing two confidence interval solutions, call them mine and yours, is for me to say that yours are too long if they are longer than mine and too short if shorter. Accordingly, define $C = \{x_1, x_2 | \bar{\theta}_{t|u} < \bar{\theta}_{\underline{x}}\}$ and regard C as a conditioning subset. Let $A_{\underline{x}}, A_{t|u}$ represent coverage of θ by the two solutions. Then $P_\theta\{A_{\underline{x}} | C\} = P_\theta\{A_{\underline{x}}\} = \gamma$ whereas $P_\theta\{A_{t|u} | C\}$ is constant over θ and $< \gamma$. The former is an indication of the "correctness" of $A_{\underline{x}}$; the latter is an example of a relevant subset in the sense of Buehler (1959) and indicates a flaw in $A_{t|u}$.

This example is of course highly special and well-behaved since we

can so easily recognize group invariance, a sufficient statistic and an exact ancillary.

7. An autoregressive example.

Consider the autoregressive model

$$x_i = \theta x_{i-1} + e_i, \quad i = 1,2,... \tag{7.1}$$

Our example requires only two observations, and writing (x,y) in place of (x_1,x_2) gives

$$\begin{aligned} x &= \theta x_o + e_1 \\ y &= \theta x + e_2 \end{aligned} \tag{7.2}$$

Assuming standard normal distributions for e_1 and e_2 we may also write

$$\begin{aligned} X &\sim N(\theta x_o, 1) \\ \{Y|X=x\} &\sim N(\theta x, 1) \end{aligned} \tag{7.3}$$

It is not known whether there exists an exactly ancillary function of X,Y. In this section we will compare some approximate ancillaries.

7.1 The predictive ancillary

The calculation of the predictive ancillary (Appendix B) begins with the fiducial distribution of θ given x:

$$\{\theta|X=x\} \sim N(x/x_o,\ 1/x_o^2) \tag{7.4}$$

Note that $x_o = 0$ is a degenerate case - x contains no information about θ when $x_o = 0$. Using (7.4) in (B.2) of Appendix B gives

$$\Psi(x,y) = \frac{x_0}{2\pi} \int_{\theta=-\infty}^{\infty} \int_{w=-\infty}^{y} e^{-Q/2} dw d\theta \qquad (7.5)$$

where

$$Q = (w - \theta x)^2 + x_0^2(\theta - x/x_0)^2.$$

Evaluation of the integral is straightforward, yielding F(u) where F is the standard normal c.d.f. and

$$u = \frac{x_0 y - x^2}{\sqrt{x_0^2 + x^2}} \quad . \qquad (7.7)$$

From the results in Appendix B we know that had we started with a translation or transformation model and if U were defined by

$$U = \frac{x_0 Y - X^2}{\sqrt{x_0^2 + X^2}} \quad , \qquad (7.8)$$

then F(U) would have a $U(0,1)$ distribution for all θ, or equivalently $U \sim N(0,1)$ for all θ. In fact the autoregressive model is neither a translation or transformation model and so we ask how strongly the distribution of U depends on θ.

It is instructive to substitute (7.2) into (7.7) to separate out the dependence on θ. This gives

$$u = \frac{-\theta x_0 e_1 + x_0 e_2 - e_1^2}{\sqrt{(1+\theta^2)x_0^2 + 2\theta x_0 e_1 + e_1^2}} \quad . \tag{7.9}$$

The last experesion purports to be approximately N(0,1) for all x_0, θ when e_1, e_2 are i.i.d. N(0,1). By inspection, the following limits behave correctly: ($\theta = \pm\infty$, x_0 arbitrary), ($x_0 = \pm\infty$, θ arbitrary). As expected $x_0 = 0$ gives problems, for in this case for any θ the distribution is negative half-normal.

7.2 Likelihood calculations and other ancillaries.

The log likelihood is

$$\begin{aligned} \ell &= -\frac{1}{2}\{(y-\theta x)^2 + (x-\theta x_0)^2\} \\ &= -\frac{1}{2} v\theta^2 + t\theta + c \end{aligned} \tag{7.10}$$

where

$$\begin{aligned} t &= x_0 x + xy \\ v &= x_0^2 + x^2 \end{aligned} \tag{7.11}$$

The Fisher information in the marginal, conditional and joint distributions is

$$i_x(\theta) = x_0^2$$

$$i_{y|x}(\theta) = x^2$$

$$i_{xy}(\theta) = x_0^2 + EX^2 = 1 + x_0^2(1+\theta^2)$$

The derivatives of the log likelihood are $\ell' = t - v\theta$ and $\ell'' = -v$ from which the maximum likelihood estimator is

$$\hat{\theta} = t/v \tag{7.13}$$

and the estimated Fisher information is

$$i_{xy}(\hat{\theta}) = 1 + x_0^2(1 + t^2/v^2). \tag{7.14}$$

The performance of the predictive ancillary (7.8) will be compared numerically with two possible competitors:

$$\begin{aligned} u_F &= \text{Fisher's ancillary} \\ &= \text{observed information} \\ &= -\ell''(\hat{\theta}) \\ &= v \\ &= x_0^2 + x^2 \end{aligned} \tag{7.15}$$

and

$$\begin{aligned} u_p &= \text{Pierce's ancillary} \\ &= (\text{observed information})/(\text{estimated information}) \\ &= v^3/\{v^2 + x_0^2(t^2 + v^2)\} \end{aligned} \tag{7.16}$$

7.3 Information in the ancillaries.

An exact ancillary by definition has a distribution free of θ. An approximate ancillary has a distribution depending only slightly on θ. One way to quantify this is by means of Fisher information. An exact ancillary is characterized by zero Fisher information. An approximate an-

cillary has small Fisher information.

The predictive ancillary U defined in (7.8) purports to be approximately standard normal and as already mentioned this is known to be exact in the limits $|\theta| \to \infty$ or $|x_o| \to \infty$. Some simulations indicated approximate normality in other cases. This suggested a crude approximation to the Fisher information based on normal approximation. Consider a family of normal distributions $U \sim N(\mu(\theta), \sigma^2(\theta))$. The information in one observation U about θ is (in a casual notation)

$$I(\theta) = I(\mu)\left(\frac{d\mu}{d\theta}\right)^2 + I(\sigma^2)\left(\frac{d\sigma^2}{d\theta}\right)^2 . \qquad (7.17)$$

(The cross term vanishes.) For the normal case $I(\mu) = \sigma^{-2}$ and $I(\sigma^2) = \frac{1}{2}\sigma^{-4}$. This gives

$$I(\theta) = \sigma^{-2}\left(\frac{d\mu}{d\theta}\right)^2 + \frac{1}{2}\sigma^{-4}\left(\frac{d\sigma^2}{d\theta}\right)^2 . \qquad (7.18)$$

The values $x_0 = \frac{1}{2}$, 1, 2 were arbitrarily chosen as intermediate between $x_0 = 0$ (awkward case) and $x_0 = \infty$ (ideal). Monte Carlo simulations were run with sample size $n = 400$ for $\theta = 0(0.20)4$ to give estimates of $\mu(\theta)$ and $\sigma^2(\theta)$, the mean and variance of U. These are given in Table 2 and graphed in Figure 1. For $x_0 = 1$, $\theta = 1$, we estimate $\sigma^2 = 0.92$, $d\mu/d\theta = 0.22$, $d\sigma^2/d\theta = -0.06$ so that from (7.18)

$$\begin{aligned} I_u(\theta) &= (0.92)^{-1}(0.22)^2 + (0.5)(0.92)^{-2}(0.06)^2 \\ &= 0.053 + 0.002 = 0.055. \end{aligned} \qquad (7.19)$$

Figure 1

Mean of Predictive Ancillary

The curves are fit by eye to $\mu(\theta)$, the mean of the predictive ancillary (7.8) for data obtained by simulation (see Appendix C). It is known that $\mu(\theta)$ tends to zero as either θ or x_0 tends to infinity.

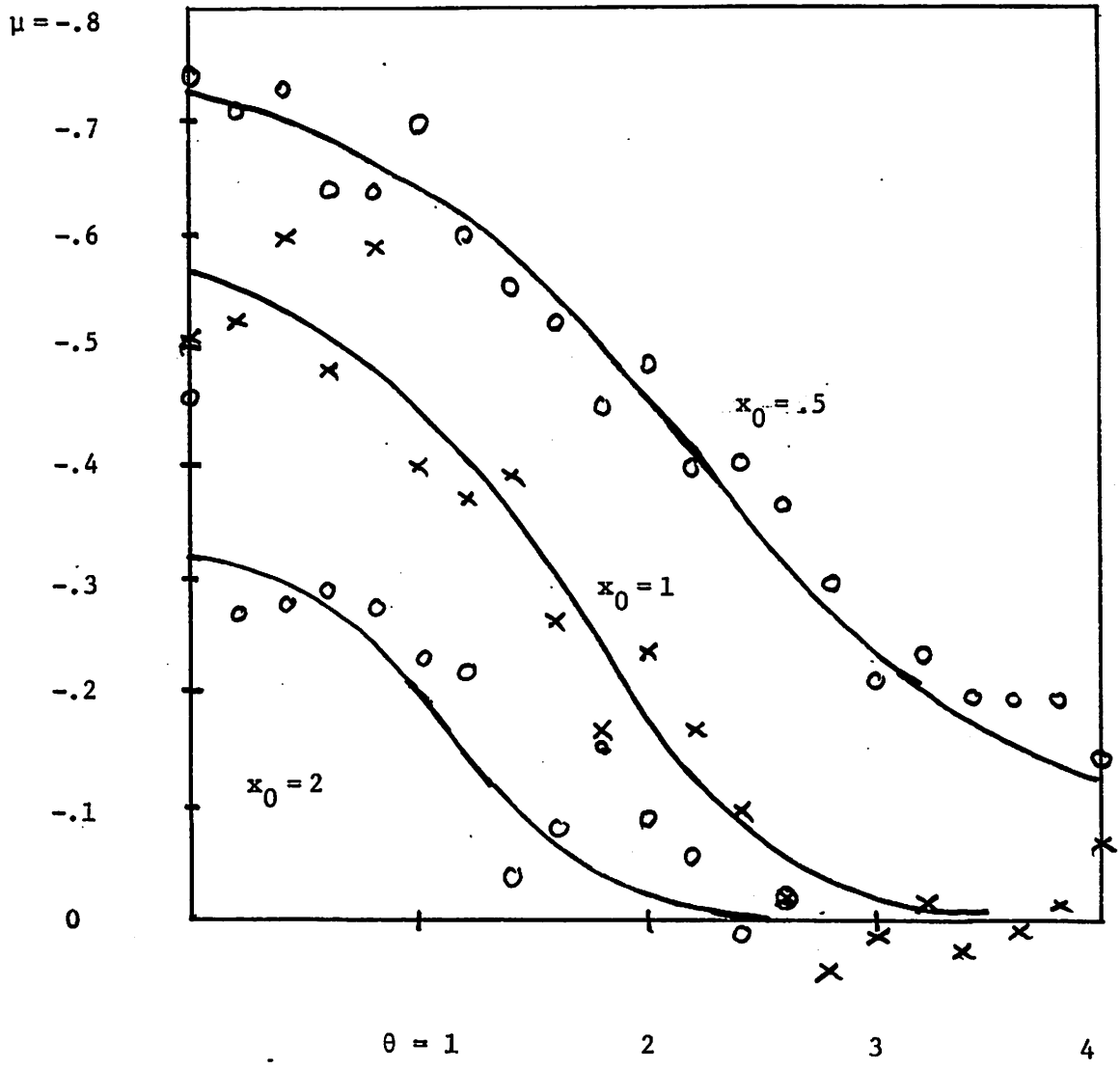

Figure 2

Variance of Predictive Ancillary

The curve is fit by eye to $\sigma^2(\theta)$, the variance of the predictive ancillary (7.8) for data obtained by simulation (see Appendix C). The curve is for $x_0 = 1$. The curves for $x_0 = 0.5$ and 2 are similar. It is known that $\sigma^2(\theta)$ tends to one as either θ or x_0 tends to infinity.

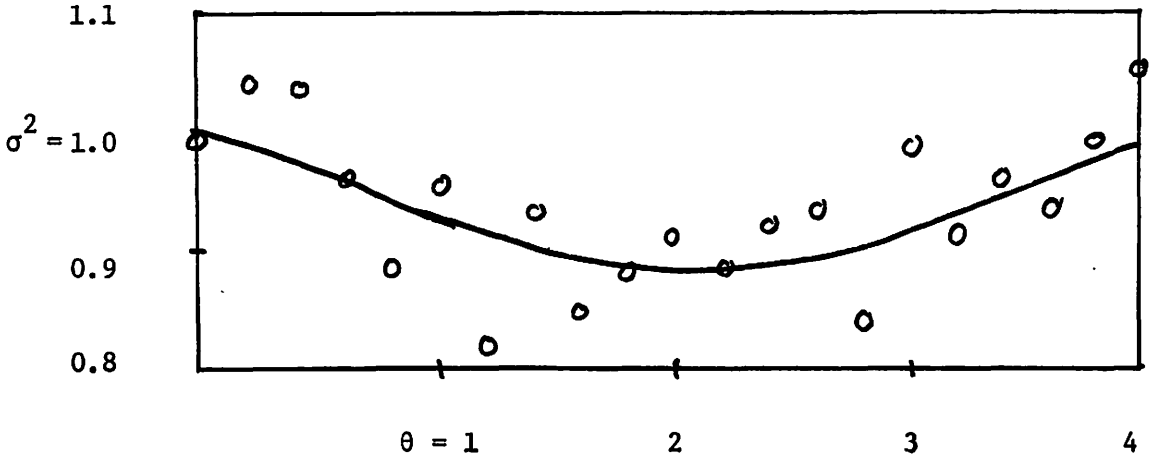

Figure 3

Mean and Variance of Pierce Ancillary

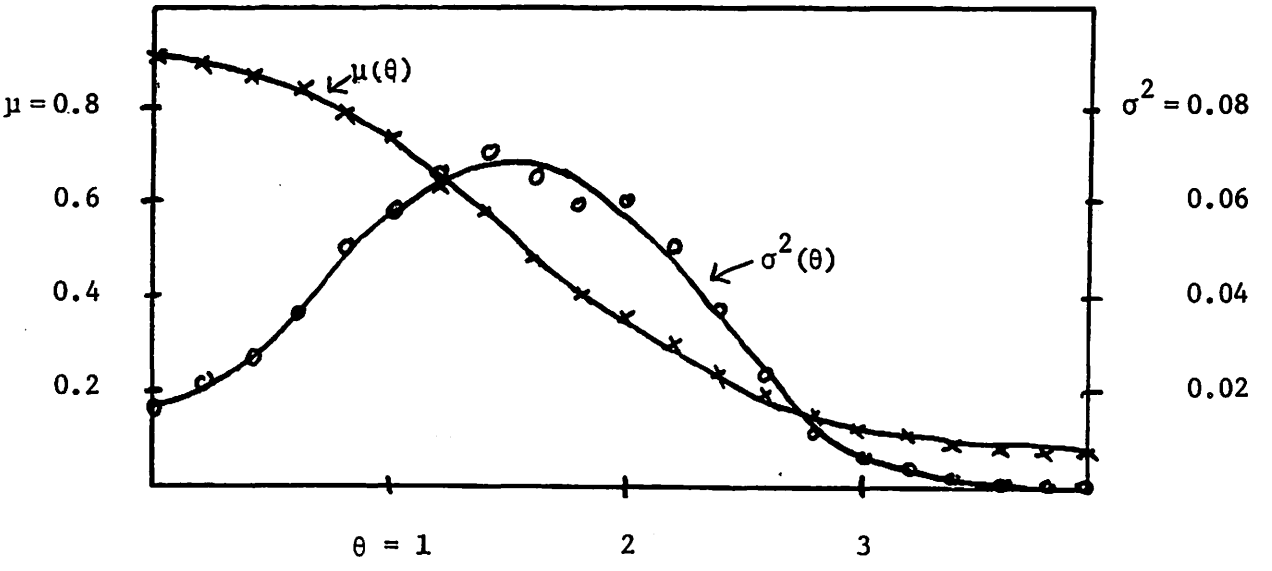

Table 2

Information in Predictive Ancillary (7.8)

x_0	θ	μ	σ^2	$\frac{d\mu}{d\theta}$	$\frac{d\sigma^2}{d\theta}$	$I_u(\theta)$	$I_{xy}(\theta)$	percent
0.5	1	-.64	.78	.14	0	.025	1.5	1.7
0.5	2	-.46	.78	.23	.04	.069	2.25	3.3
0.5	3	-.22	.82	.15	.05	.030	3.25	.9
0.5	4	-.15	.86	.08	.04	.008	5.25	0
1	1	-.46	.92	.22	-.06	.055	3	1.8
1	2	-.18	.88	.32	0	.114	6	1.9
1	3	-.02	.92	.06	.06	.006	11	0
1	4	0	.97	0	.03	.001	18	0
2	1	-.20	1.02	.18	-.06	.033	9	.3
2	2	-.05	.96	.10	-.03	.011	21	0
2	3	0	.94	0	0	.001	41	0
2	4	0	.96	0	0	0	69	0

Here μ, σ^2, $d\mu/d\theta$, $d\sigma^2/d\theta$ are estimated from the plots in Figure 1 which were fitted by eye. $I_u(\theta)$ and $I_{xy}(\theta)$ are calculated from (7.18) and (7.12) respectively. The last column shows $I_u(\theta)$ as a percent of $I_{xy}(\theta)$. These rather crude calculations indicate that for $x_0 > 0.5$, and for all θ, the predictive ancillary u contains less than four percent of the Fisher information in (x,y).

This value can be compared with the Fisher information in X,Y, which by (7.12) is $1+x_0^2\ (1+\theta^2) = 3$. For $x_0 > 0.5$ and for all θ we estimate from Table 2 that the predictive ancillary (7.8) contains less than four percent of the total Fisher information. The mean of U contributes much more (in (7.18)) than the variance.

For comparison we did a similar analysis on the Pierce ancillary. For $x_0 = 1$, $\theta = 1$, we estimate $\sigma^2 = 0.06$, $d\mu/d\theta = -0.3$, $d\sigma^2/d\theta = 0.05$ (see Figure 3), giving

$$\begin{aligned} I(\theta) &= (0.06)^{-1}(0.3)^2 + (0.5)(0.06)^{-2}(0.05)^2 \\ &= 1.5 + 0.35 = 1.85, \end{aligned} \tag{7.20}$$

which shows that Pierce's ancillary may contain about sixty percent of the total information. This should however be considered a very crude approximation since our simulations showed a highly skewed and therefore nonnormal distribution.

I should point out that no criticism of Pierce is intended. His statistic (which was incidentally named by me, Buehler, 1982, p. 593), performs beautifully elsewhere, in particular in Appendix A, Example A5. If we were to draw a moral at this point it would be that we should test the appropriateness of any ancillary in each application and not expect a single formula to work every time.

Comparison can also be made with Fisher's ancillary, which by (7.15) is $u_F = x_0^2 + x^2$. The constant x_0^2 can be ignored for purposes of calculating the information in u_F. Putting $u_F^* = x^2$ we see that u_F^* has a noncentral chi square distribution with one degree of freedom.

The Fisher information in Fisher's ancillary probably cannot be put in closed form. We may observe that by (7.12) the information in x is x_0^2, or 1 when $x_0 = 1$, and the information in x^2 is slightly less than the information in x. By this estimate, when $x_0 = \theta = 1$, Fisher's ancillary, $x_0^2 + x^2$ contains slightly less than one third of the information in the bivariate observation (x,y). The normal approximation was found to give a higher estimate presumably owing to the skewness of the distribution of x^2.

For the record we will give some algebraic expressions leading to the Efron-Hinkley ancillary even though it was not studied by simulations.

For t,v defined by (7.11) we have

$$\begin{aligned} Et &= \theta\{1 + x_0^2(1+\theta^2)\} \\ Ev &= 1 + x_0^2(1+\theta^2) \\ \text{Var } t &= 1 + 2\theta^2 + x_0^2(1 + 5\theta^2 + 4\theta^4) \\ \text{Var } v &= 2 + 4x_0^2\theta^2 \\ \text{Cov}(t,v) &= 2\theta\{1 + x_0^2(1+2\theta^2)\} \quad . \end{aligned} \tag{7.21}$$

The second order moments of (ℓ', ℓ'') are (from (7.10) and 7.21))

$$\begin{aligned} \sigma_1^2 &= \text{Var } \ell' = 1 + x_0^2(1+\theta^2) \\ \sigma_2^2 &= \text{Var } \ell'' = 2 + 4x_0^2\theta^2 \\ \sigma_{12} &= \text{Cov}(\ell', \ell'') = -2x_0^2\theta \ . \end{aligned} \tag{7.22}$$

The Efron curvature is (from (A.22) and (7.22))

$$\gamma(\theta) = \{\sigma_1^2\sigma_2^2 - \sigma_{12}^2\}^{1/2}/\sigma_1^3$$

$$= 2^{1/2}\{1+x_0^2(1+3\theta^2) + x_0^4\theta^2\}^{1/2}\{1 + x_0^2(1+\theta^2)\}^{-3/2}. \tag{7.23}$$

The Efron-Hinkley ancillary (or affine ancillary) is by (A.25)

$$u_{EH} = (1-u_p)/\hat{\gamma} \tag{7.24}$$

where u_p is given by (7.16) and $\hat{\gamma} = \gamma(\hat{\theta})$ is obtained from (7.23) by substituting $\theta = \hat{\theta} = t/v$.

8. Discussion and Summary.

A confidence distribution can be obtained as an induced distribution using the conditional distribution of t given u. Here t is typically the MLE $\hat{\theta}$, while u an ancillary statistic. Ideally u is exactly ancillary (distribution constant), but we have also considered the case where u is only approximately ancillary. Choosing u and finding the conditional distribution of $\hat{\theta}$ given u has been called a "conditionality resolution" by Barndorff-Nielsen. We have discussed in Section 4 some criteria which might be used to define optimal conditionality resolutions.

Efron and Hinkley (1978), following Fisher, have suggested using $u = \hat{\ell}''$, the second derivative of the log likelihood evaluated at the maximum, which we call Fisher's ancillary. An approximate conditionality resolution suggested by Efron and Hinkley involves the approximation Var $(\hat{\theta}|\hat{\ell}'') \approx 1/\hat{\ell}''$.

But $\hat{\ell}''$ need not be approximately ancillary since in extreme cases it can be a sufficient statistic itself (Barndorff-Nielsen, 1978). A sufficient statistic contains all of the Fisher information in the sample, while an exact ancillary contains zero Fisher information. We have thus suggested using Fisher information as a criterion for judging the suitability of approximate ancillaries.

In Appendix A we have defined translation and transformation models and have shown how they relate to the following ancillaries: likelihood shape, Fisher's, Pierce's and Efron and Hinkley's. It is proved that for translation families, the likelihood shape ancillary solves Fisher's Problem of the Nile.

In Appendix B we introduced a new ancillary which we have called (for want of a better term) the predictive ancillary. It is shown to be exact (distribution constant) for translation and transformation models.

The autoregressive model considered in Section 7 furnishes an example in which no exact ancillary is known, but the predictive ancillary can be expressed in simple closed form. Monte Carlo simulations show that it contains only a small percentage of the Fisher information over a range of parameter values.

Acceptance of an ancillary such as the predictive ancillary in a model such as the autoregressive model only partially solves the problem of conditionality resolution because it remains to find the conditional distribution of $\hat{\theta}$ given the ancillary. That distribution would furnish conditional confidence limits. These last steps seemingly require rather extensive simulations, and these have not been attempted for the present

paper. The Efron-Hinkley approach, beginning with Fisher's ancillary, when applicable, has the advantage of supplying directly from the observed likelihood function both the ancillary and the approximate conditional variance.

Appendix A

Translation and Transformation Models

A1. Translation models.

The MLE of θ will be denoted by $\hat{\theta}$. Regularity conditions such as existence, uniqueness, smoothness are assumed throughout, but are not explicitly stated.

Definition A1. $f(x;\theta)$ is a translation model if x is one-to-one with $(\hat{\theta},u,v)$ where $(\hat{\theta},u)$ is sufficient for θ, u is distribution constant (exactly ancillary), and the conditional density $f(\hat{\theta}|u;\theta)$ has the form of $f_o(\hat{\theta}-\theta|u)$.

Lemmas A1 through A3 were asserted without proof in Buehler (1982, p. 593).

Lemma A1. If $f(x;\theta)$ is a transformation model then $f(x;\theta) \propto f_o(\hat{\theta}-\theta|u)$.

Proof. If J is the Jacobian of the transformation from x to $(\hat{\theta},u,v)$, then using sufficiency and ancillarity we have

$$\begin{aligned} f(x;\theta) &= f(\hat{\theta},u,v;\theta)J \\ &= f(\hat{\theta},u;\theta)f(v|\hat{\theta},u)J \\ &= f(\hat{\theta}|u;\theta)f(u)f(v|\hat{\theta},u)J \\ &\propto f(\hat{\theta}-\theta|u) \end{aligned}$$

If x_1 and x_2 are values of x then we will denote $\hat{\theta}(x_j)$ by $\hat{\theta}_j$ and similarly for other statistics.

Lemma A2. If $f(x;\theta)$ is a translation model and $w(x)$ satisfies

$$w_1 = w_2 \Leftrightarrow f(x_1;\theta+\hat{\theta}_1) \propto f(x_2;\theta+\hat{\theta}_2),$$

then: (i) $(\hat{\theta},w)$ is minimal sufficient, (ii) $u_1 = u_2$ implies $w_1 = w_2$ (w is a

function of (u), (iii) w is distribution constant.

Proof. (i) Using the known minimal sufficiency of the (normalized) likelihood it suffices to show $(\hat{\theta}_1,w_1) = (\hat{\theta}_2,w_2) \Leftrightarrow f(x_1;\theta) \propto f(x_2;\theta)$. Assume the latter. Then $\hat{\theta}_1 = \hat{\theta}_2$ and $f(x_1;\theta+\hat{\theta}_1) \propto f(x_2;\theta+\hat{\theta}_1) = f(x_2;\theta+\hat{\theta}_2)$, and so $w_1 = w_2$. Next assume $(\hat{\theta}_1,w_1) = (\hat{\theta}_2,w_2)$.

$$f(x_1;\theta') = f(x_1;\theta+\hat{\theta}_1) \quad \text{where } \theta' = \theta+\hat{\theta}_1$$

$$\propto f(x_2;\theta+\hat{\theta}_2) \quad \text{because } w_1 = w_2$$

$$= f(x_2;\theta+\hat{\theta}_1) \quad \text{because } \hat{\theta}_1 = \hat{\theta}_2$$

$$= f(x_2;\theta') \quad \text{for all } \theta'.$$

(ii) Assume $u_1 = u_2$. Then

$$f(x_1;\theta+\hat{\theta}_1) \propto f_0\left(\hat{\theta}_1 - (\theta+\hat{\theta}_1);u_1\right) \quad \text{by Lemma A1}$$

$$= f_0(-\theta;u_1)$$

$$= f_0(-\theta;u_2) \quad \text{since } u_1 = u_2$$

$$= f_0\left(\hat{\theta}_2 - (\theta+\hat{\theta}_2);u_2\right)$$

$$\propto f\ (x_2;\theta+\hat{\theta}_2) \quad \text{by Lemma A1,}$$

which shows that $w_1 = w_2$.

(iii) This is a consequence of (ii).

Remark. Lemma A2 states that if we have a translation model, then the likelihood function determines an exact ancillary statistic w such that

$(\hat{\theta},w)$ is minimal sufficient. The statistic w simply specifies the shape of the likelihood apart from its location $\hat{\theta}$. The shape w and location $\hat{\theta}$ together determine the likelihood function.

Example A1. If $f(x;\theta) = \prod f(x_i - \theta)$ and there is no sufficiency reduction beyond the order statistic, then w is the order statistic spacings $(x_{(2)} - x_{(1)},\ldots,x_{(n)} - x_{(1)})$, which Fisher called the "complexion" of the sample.

Example A2. If $f(x;\theta) = f(x_1 - \theta,\ldots,x_n - \theta)$ and there is no sufficiency reduction then $w = (x_2 - x_1,\ldots,x_n - x_1)$.

Example A3. (Efron and Hinkley, 1978, p. 458). The parameter θ is measured by one of two measuring instruments whose errors are $N(0,\sigma_1^2)$, $N(0,\sigma_2^2)$. If it is decided at random which instrument is used for each of n measurements, then w gives the number of times the first instrument is used.

Example A4. (Counterexample) Likelihood shape is not always ancillary. Consider the autoregressive model

$$x_j = x_{j-1} + e_j \qquad j = 1, 2,\ldots, n \qquad (A.1)$$

where e_j are i.i.d. $N(0,1)$. The log likelihood equals a constant plus $t\theta - v\theta^2/2$ where

$$t = \sum_{i=1}^{n} x_i x_{i-1} \quad \text{and} \quad v = \sum_{i=1}^{n} x_{i-1}^2 \qquad (A.2)$$

Then $\hat{\theta} = t/v$, $(\hat{\theta},v)$ is sufficient, and the likelihood shape, apart from location, is determined by v. For $n=2$, $x_o = 1$, $Ev = E(x_o^2 + x_1^2) = 2 + \theta^2$, showing that v is not distribution constant. For more about this example see Section 7.

Define the log likelihood function by full and abbreviated (x suppressed) notations by

$$\ell(\theta;x) = \ell_\theta = \log f(x;\theta).$$

Derivatives with respect to θ will be denoted with varying degrees of abbreviation as

$$\begin{aligned} \partial\ell(\theta;x)/\partial\theta &= \ell'(\theta;x) = \ell'_\theta = \ell' \\ \partial^2\ell(\theta;x)/\partial\theta^2 &= \ell''(\theta;x) = \ell''_\theta = \ell'' \\ \partial^n\ell(\theta;x)/\partial\theta^n &\quad \ell^{(n)}(\theta;x) = \ell^{(n)}_\theta = \ell^{(n)} \end{aligned} \tag{A.3}$$

Values at the maximum will be further abbreviated for example by

$$\hat{\ell}'' = \ell''_{\hat\theta} = \ell''(\hat\theta;x). \tag{A.4}$$

Then $\hat{\ell}' = 0$ (regularity always assumed). We will call

$$I(x) = -\hat{\ell}'' \tag{A.5}$$

the observed Fisher information, and we will call

$$i(\theta) = E_\theta(\ell')^2 = -E_\theta\ell'' \tag{A.6}$$

the (expected) Fisher information, and we will call

$$\hat{i} = \hat{i}(x) = i(\hat\theta) \tag{A.7}$$

the estimated Fisher information.

Lemma A3. To the extent that $\ell(\theta;x)$ is an analytic function in θ with a convergent expansion about $\hat\theta$, the statistic w of Lemma 2 is equivalent to $\tilde{w} = (\hat{\ell}'', \hat{\ell}''', \ldots)$

Proof. Both statistics determine the shape of ℓ_θ about $\theta = \hat\theta$ and are determined by that shape.

Remark. While $(\hat\theta, \tilde{w})$ is in fact minimal sufficient, $\tilde{w}$ nevertheless

contains redundancies in that a finite vector will suffice to determine the rest. For either Example 1 or Example 2 we could reasonably expect to use $(\hat{\ell}'', \ldots, \hat{\ell}^{(n)})$. For Example 3, $\hat{\ell}''$ suffices. Examples could easily be constructed in which $\hat{\ell}''$ is constant and w is equivalent to $\hat{\ell}'''$ (two measuring instruments with equal Fisher information but different skewness).

An interesting question is whether ancillarity of $\tilde{w}$ and in particular of $\hat{\ell}'' = I(x)$ carries over at least approximately to other models, and if so, how this should affect our inferences. These questions have been addressed by Fisher (1925, 1934) and by Efron and Hinkley (1978). Intuitively a large value of $I(x)$ corresponds to a sharply peaked likelihood and so, for a Bayesian, to a sharply peaked posterior. Thus large $I(x)$ seemingly indicates high precision.

Barndorff-Nielsen (1978) pointed out a basic flaw in the theory that $I(x)$ is approximately ancillary: Ancillarity of $I(x)$ depends on the parametrization. That is, $I(x)$ can be exactly ancillary in one parametrization, but far from exact in another. We explore this further in the next section.

Barnard and Sprott (1971) used the likelihood shape to choose between competing ancillaries. But shapes change with reparametrization, so that the definition of a shape statistic is a tricky problem.

A2. Transformation models.

Definition A2. A model $f(x;\theta)$ will be called a transformation model if there exists a smooth one-to-one transformation $\tau = \psi(\theta)$ such that the transformed model $f(x;\tau)$ is a translation model. (Then $f(\hat{\tau}|u;\tau) = f_o(\hat{\tau}-\tau|u)$, etc.).

Definition A3. For any model $f(x;\theta)$, $\tau = \psi(\theta)$ will be called a canonical parameter if $(d\tau/d\theta)^2 = i_\theta$, the expected Fisher information. (The Fisher information with respect to τ, i_τ, is then unity.)

A canonical parameter is unique up to change of sign and additive constant. For any transformation model we can in principle find the canonical parameter by finding the Fisher information i_θ and taking the indefinite integral of $i_\theta^{1/2}$. This "variance stabilizing" transformation appears often in the literature; see for example Efron and Hinkley (1978).

Now suppose we are faced with a transformation model, or an approximate or suspected transformation model, how can we find an exact or approximate ancillary? One way is to find the Fisher information i_θ, change to the canonical parameter, and appeal to Lemma A3. In principle this is guaranteed (in sufficiently regular cases) to unmask any transformation model and yield an ancillary in such that $(\hat{\theta}, u)$ is minimal sufficient. There is incidentally an alternative way to find the canonical parameter for any transformation model. Let $F(\hat{\theta}|\theta)$ be the c.d.f. of $\hat{\theta}$. Then the following factorization will be possible:

$$\frac{\partial F(\hat{\theta}|\theta)/\partial\hat{\theta}}{\partial F(\hat{\theta}|\theta)/\partial\theta} = \frac{a(\hat{\theta})}{b(\theta)}, \tag{A.8}$$

and the canonical parameter will satisfy $d\tau/d\theta \propto b(\theta)$.

Let ℓ_0 denote likelihood as a function of the canonical parameter τ and ℓ denote likelihood as a function of the original parameter θ. Then

$$\ell(\theta) = \ell_0(\tau(\theta)), \tag{A.9}$$

and differentiating twice and setting $\theta = \hat{\theta}$ gives

$$\hat{\ell}'' = \hat{\ell}_0''(d\hat{\tau}/d\hat{\theta})^2 = \hat{\ell}_0''\hat{i} \tag{A.10}$$

where $\hat{i}$ is the maximum likelihood estimator of i_θ and hence of $(d\tau/d\theta)^2$.

This expression, previously noted by Efron and Hinkley (1978), eq. (2.14), is the observed information analog of the well known identify $i_\theta = i_\tau (d\tau/d\theta)^2$.

If the given model is a transformation model, then τ is a translation parameter and $\hat{\ell}_0''$ is exactly ancillary. Thus in terms of the original parameter θ the statistic

$$u_p = -\hat{\ell}''/\hat{i} \tag{A.11}$$

is exactly ancillary. It is of interest to ask whether it is approximately ancillary in other models. In words u_p can be simply described as the observed information divided by the estimated expected information. Using a different argument, Pierce (1975) arrived at

$$u_p - 1 = (-\hat{\ell}'' - \hat{i})/\hat{i} \tag{A.12}$$

as an approximate ancillary statistic. Accordingly I have suggested (Buehler, 1982, p. 593) that u_p be called Pierce's ancillary. It is reasonable to suppose u_p is usually superior to Fishers $\hat{\ell}''$ because it is transformation invariant and exact for transformation models.

Taking the k-th derivative of (A.9) and putting $\theta = \hat{\theta}$ yields a k-th order Pierce ancillary. This sequence of statistics has previously been described (in a different notation) by Barndorff-Nielsen (1982).

Of course the number of nontrivial statistics in the sequence corresponds to the number of constants needed to determine the likelihood shape. The second order statistic incidentally works out to

$$\ell_0''' = \ell'''\hat{i}^{-3/2} - (3/2)\hat{\ell}''\hat{i}'\hat{i}^{-5/2}, \tag{A.13}$$

which is obtainable from

$$\ell''' = \ell_0'''\tau'^3 + 3\ell_0''\tau'\tau + \ell_0'\tau''' ,$$
$$\tau' = i^{1/2}, \text{ and } \tau'' = (1/2)i^{-1/2}i' . \tag{A.14}$$

Example A5. Normal with known coefficient of variation (Hinkley, 1977), $X \sim N(\theta, c^2\theta^2)$. We restrict $\theta > 0$ and for convenience take $c = 1$. We find

$$\ell_\theta = -n\log\theta - \tfrac{1}{2}nS_2\theta^{-2} + nS_1\theta^{-1} \tag{A.15}$$

where $S_j = (1/n)\Sigma x_i^j$, and

$$\ell_\theta' = n\theta^{-3}(-\theta^2 - S_1\theta + S_2) \tag{A.16}$$

from which

$$\hat{\theta} = \tfrac{1}{2}S_1\{(1+4r)^{1/2} - 1\} \qquad \text{where } r = S_2/S_1^2 \tag{A.17}$$

Further

$$\ell_\theta = n\theta^{-4}\{\theta^2 + 2S_1\theta - 3S_2\} \tag{A.18}$$

from which

$$i_\theta = -E_\theta\ell_\theta'' = k\theta^{-2} \qquad \text{for some } k \tag{A.19}$$

and

$$u_P = \hat{\ell}''/\hat{i} = (n/k)\hat{\theta}^{-2}\{\hat{\theta}^{-2} + 2S_1\hat{\theta} - 3S_2\}$$
$$= (n/k)\ \{1 + 2\varphi^{-1} - 3r\varphi^{-2}\} \tag{A.20}$$

where

$$\varphi = \varphi(r) = \tfrac{1}{2}\{(1+4r)^{1/2} - 1\} \tag{A.21}$$

Thus u_P is a function of r, and the calculation leads to the discovery of the ancillary r. Of course the main features of this model are clear from sufficiency and invariance considerations. To complete the analysis

one routinely shows that r is in fact ancillary and that $(\hat{\theta},r)$ is minimal sufficient (as expected since the sufficient statistic (S_1, S_2) is two dimensional).

One way to refine the Pierce statistic is to consider the joint distribution of ℓ' and ℓ''. We have $E_\theta \ell' = 0$, $E_\theta \ell'' = -i_\theta$. Define $\sigma_1^2 = \text{Var } \ell' = i_\theta$, $\sigma_2^2 = \text{Var } \ell''$, $\sigma_{12} = \text{Cov}(\ell', \ell'')$, $\rho = \sigma_{12}/\sigma_1\sigma_2$. The Efron curvature is defined by (Efron, 1975)

$$\gamma = (\sigma_2/\sigma_1^2)\sqrt{1-\rho^2} \tag{A.22}$$

Suppose we form a linear function $w = c_0 + c_1\ell' + c_2\ell''$ where c_0, c_1, c_2 may depend on θ but not on x. Determine these constants by the three conditions

$$Ew = 0, \quad \text{Var } w = 1, \quad \text{Cov}(w, \ell') = 0 \tag{A.23}$$

The first two conditions are directed at creating an approximate ancillary (ancillary in mean and variance), while the third seeks orthogonality. The resulting w depends on θ, but we substitute $\hat{\theta}$ to obtain

$$u_{EH} = w(\hat{\theta}) = \hat{c}_0 + \hat{c}_1\hat{\ell}' + \hat{c}_2\hat{\ell}'' = \hat{c}_0 + \hat{c}_2\hat{\ell}'' \tag{A.24}$$

A little algebra shows that this reduces to plus or minus

$$u_{EH} = (1 - u_p)/\hat{\gamma} = (\hat{\ell}'' + \hat{i})/(\hat{i}\hat{\gamma}) \tag{A.25}$$

which Efron and Hinkley (1978) called $Q(x)$. The above derivation follows Barndorff-Nielsen (1980).

Since it is known that γ is transformation invariant and constant for translation models (Efron, 1975) it follows that u_{EH} is equivalent to u_p for transformation models and is therefore exactly ancillary for transformation models.

Barndorff-Nielsen (1980) has generalized u_{EH} by bringing in more derivatives (ℓ''', etc.) and by considering vector parameters. His derivation is motivated by theory of exponential families, and his so-called affine ancillaries are shown to be exactly distribution constant in certain nontranslation models.

The statistic u_{EH} has also been studied by Peers (1978).

Appendix B

An Ancillary Based on Predictive Distributions

In this appendix we show how to construct a statistic which is distribution constant (exactly ancillary) for translation and transformation models and presumably approximately ancillary for other models.

Let $X \in R^n$ denote a vector of past observations and let $Y \in R^1$ denote a single future observation. Suppose X, Y have joint distribution P_θ, $\theta \in \Omega$. Either Bayesian or non-Bayesian methods can be used to obtain a "predictive distribution" of Y given X. (A good general reference is Aitchison and Dunsmore (1975).) By a (non-Bayesian) predictive distribution we mean a set of upper prediction limits $L(x,\gamma)$, defined for $0 < \gamma < 1$ and satisfying

$$P_\theta\{Y \leq L(X;\gamma)\} = \gamma \quad \text{for all } \theta \in \Omega \ . \tag{B.1}$$

Thus $L(x;\gamma)$ is interpreted as the γ percentile of the predictive distribution of Y given that $X = x$.

If we define an ancillary statistic to be a function of (X,Y) whose distribution is the same for all θ, then finding a function L satisfying (B.1) is essentially equivalent to finding an ancillary statistic. This is stated more formally below as Proposition B1. In the following Lemma think of Ψ as the c.d.f. of some Y and $L(\gamma)$ as the γ percentile.

Lemma B1. If $\Psi(y)$ is continuous and nondecreasing for $-\infty < y < \infty$, $0 \leq \Psi(y) \leq 1$, and $L(\gamma) = \sup\{y | \Psi(y) \leq \gamma\}$, then $L(\gamma)$ is strictly increasing and right-continuous, and $\{y|\Psi(y) \leq \gamma\} = \{y|y \leq L(\gamma)\}$ for all $0 \leq \gamma \leq 1$. Contrariwise if $L(\gamma)$ with the stated properties is given, then put $\Psi(y) = \inf\{\gamma | y \leq L(\gamma)\}$,

and the same relationships hold.

Note that if Ψ is strictly increasing, then L is simply the inverse function. The lemma just shows how to handle intervals where Ψ is constant. In the following proposition think of $\Phi(x,y)$ as the c.d.f. of the predictive distribution of Y given $X = x$.

Proposition B1. Let (X,Y) have distribution P_θ, $\theta \in \Omega$. The following three conditions are equivalent: (i) There exists a real-valued function $\Phi(x,y)$ which is nondecreasing in y for each x such that the random variable $\Phi(X,Y)$ has a continuous distribution which is the same for all $\theta \in \Omega$. (ii) There exists a real-valued function $\Psi(x,y)$ such that $\Psi(X,Y)$ is uniformly distributed on $(0,1)$ for all $\theta \in \Omega$. (iii) There exists a real-valued function $L(x;\gamma)$, strictly increasing and right continuous for γ for each x, which satisfies (B.1).

Proof. Assume (i). Put $\Psi(x,y) = F(\Phi(x,y))$, where F is the c.d.f. of Φ. Then Ψ satisfies (ii). Assume (ii). Then $\Phi = \Psi$ satisfies (i). Thus (i) iff (ii). Assume (ii). Put $L(x;\gamma) = \sup\{y | \Psi(x,y) \leq \gamma\}$. By Lemma B1, for each x,

$$\{y | \Psi(x,y) \leq \gamma\} = \{y | y \leq L(x;\gamma)\}$$

and

$$P_\theta\{Y \leq L(X;\gamma)\} = P_\theta\{\Psi(X,Y) \leq \gamma\} = \gamma$$

so that (iii) is satisfied. Assume (iii). Define $\Psi(x,y) = \inf\{\gamma | y \leq L(x;\gamma)\}$ and a similar application of Lemma B1 gives (ii).

The relationship in Proposition B1 can be used to find ancillary statistics from predictive distributions or vice versa. Although our main

interest at the moment is the possibility of finding approximate ancillaries from approximate predictive distributions, we first mention a few exact results.

Formally we can write

$$\Psi(x,y) = P\{Y \leq y|x\} = \int P\{Y \leq y|x;\theta\}f(\theta|x)d\theta \ . \qquad (B.2)$$

In Bayesian analysis, $f(\theta|x)$ is a posterior density. In non-Bayesian analysis we may try anything we like and check the consequences.

When $f(\theta|x)$ is a posterior corresponding to prior $\pi(\theta)$ we will say $\Psi(x,y)$ is the "Bayes (predictive) ancillary" (relative to π). There is no reason to suppose Ψ would be exactly ancillary in this case and one would have to check whether it were approximately ancillary in specific cases.

When $f(x|\theta)$ is a fiducial distribution, then we will say $\Psi(x,y)$ in (B.2) is the "(Fiducial) predictive ancillary." That this yields an exact ancillary in certain cases is shown in Theorem B1 and Corollary B1 and B2. A non-exact case is considered in Section 7 (the autoregressive model).

Theorem B1. Let F_U and F_V be c.d.f's of an absolutely continuous variate U and any variate V; let $X = U + \theta$, $Y = V + \theta$, and

$$\Psi(x,y) = \int F_V(y-\theta)f_U(x-\theta)d\theta \qquad (B.3)$$

where $f_U = F_U'$. Then $\Psi(X,Y) \sim U(0,1)$.

Proof.

$$\Psi(x,y) = \int_{-\infty}^{\infty} F_V(y-x+u)f_U(u)du$$

$$= \int_{-\infty}^{\infty} f_U(u) \int_{\infty}^{y-x+u} dF_V(v)du$$

$$= P\{V - U \leq y - x\} \qquad (B.4)$$

$$= P\{Y - X \leq y - x\}$$

Thus $\Psi(x,y) = F_Z(y - x)$ where $Z = Y - X = V - U$. Absolute continuity of U implies absolute continuity of Z, and $\Psi(X,Y) = F_Z(Z) \sim U(0,1)$.

Notice that (B.2) and (B.3) correspond with $f(\theta|x) = f_U(x - \theta)$, the fiducial distribution of θ given $X = x$. Proposition B2 actually follows from Theorem 3 of Hora and Buehler (1967). Their result shows how to get predictive distributions in more general invariant models.

Corollary B1. If $f(x;\theta)$ is a translation model such that $\hat{\theta}$ has conditional density $f_0(\hat{\theta} - \theta|u)$ as described in Definition A1 (Appendix A), V is any real valued variable, $T = V + \theta$, and

$$\Psi(t,u,y) = \int F_V(y - \theta) f_0(t - \theta|u)d\theta, \qquad (B.5)$$

then the random variable $\Psi(T,U,Y)$ (where T given u has the distribution of $\hat{\theta}$ given u) has a $U(0,1)$ distribution for all θ.

Proof. By Theorem B1 the result holds conditionally for each $U - u$; hence it holds for any distribution of U.

Theorem B1 which applies to translation models can be restated for transformation models. If X, Y are replaced by W, Z which follow transformation models then in (B.2) we would formally substitute the fiducial density of θ given w in the form $-\partial F(w;\theta)/\partial\theta$ for $f(\theta|w)$.

Corollary B2. Suppose W, Z are independent with c.d.f.'s $F_1(w;\theta)$

and $F_2(z;\theta)$, and suppose there exist smooth increasing transformations $x=\phi_1(w)$, $y=\phi_2(z)$, $\tau=\psi(\theta)$ such that τ is a location parameter for both $X=\phi_1(W)$ and $Y=\phi_2(Z)$. Define

$$\Psi(w,z) = -\int F_2(z;\theta)\,\frac{\partial F_1(w;\theta)}{\partial\theta}\,d\theta\ .$$

Then $\Psi(W,Z)$ has a $U(0,1)$ distribution for all θ.

Proof. Define $F_3(x-\tau) = P\{X\le x;\tau\}$ and $F_4(y-\tau) = P\{Y\le y;\tau\}$.

Then $F_2(z;\theta) = P\{Z\le z;\theta\} = P\{Y\le y;\theta\} = F_4(y-\psi(\theta))$. Similarly $F_1(w;\theta) = F_3(x-\psi(\theta))$, and $\partial F_1(w,\theta)/\partial\theta = (\partial F_3(x-\psi(\theta))/\partial\psi)d\psi/d\theta = -f_3(x-\psi(\theta))d\psi/d\theta$ where $f_3(x) = dF_3(x)/dx$. Substituting in (B.6) and changing the variable of integration from θ to $\tau=\psi(\theta)$ transforms (B.6) into the form (B.3).

Appendix C

Monte Carlo Simulations

Table C.1 gives the mean $\mu(\theta)$ and variance $\sigma^2(\theta)$ of the distribution of the predictive ancillary (7.8) as approximated by simulation. Each tabled value is based on a sample of size $n = 400$. Normal random variables were obtained by summing 12 uniform random variables provided by the computer (Atari 800). By visual inspection the distributions of the predictive ancillary appeared to be approximately normal, but only the mean and variance were recorded.

Table C.2 gives similar results for the Pierce ancillary (7.16). These distributions appeared to be highly skewed. Here again $n = 400$.

Table C.1

Mean and Variance of Predictive Ancillary

θ	x_0	$\mu(\theta)$	$\sigma^2(\theta)$
0	.5	-.738205	.798399
.2	.5	-.711956	.795691
.4	.5	-.728561	.838237
.6	.5	-.639	.761768
.8	.5	-.641994	.794786
1	.5	-.698618	.736389
1.2	.5	-.599284	.81987
1.4	.5	-.547651	.820749
1.6	.5	-.517571	.845203
1.8	.5	-.447372	.773949
2	.5	-.487499	.771028
2.2	.5	-.395688	.882992
2.4	.5	-.400509	.835125
2.6	.5	-.364527	.881238
2.8	.5	-.296647	.933833
3	.5	-.207569	.825686
3.2	.5	-.230177	.860128
3.4	.5	-.193071	.732316
3.6	.5	-.193768	.816191
3.8	.5	-.195378	.902088
4	.5	-.138075	.841261
-3.72529E-09	1	-.508856	.996589
.2	1	-.525613	1.04121
.4	1	-.597625	1.03753
.6	1	-.484583	.96128
.8	1	-.593451	.884573
1	1	-.398467	.954954
1.2	1	-.369603	.817963
1.4	1	-.392538	.932391
1.6	1	-.261953	.84351
1.8	1	-.162815	.880843
2	1	-.233791	.908388
2.2	1	-.165297	.885722
2.4	1	-9.41107E-02	.919802
2.6	1	-1.48824E-02	.934011
2.8	1	4.66988E-02	.834966
3	1	1.54975E-02	.985471
3.2	1	-.134074	.908828
3.4	1	3.02107E-02	.961806
3.6	1	1.75602E-02	.933261
3.8	1	-9.90801E-03	.987654
4	1	6.88485E-02	1.05017

Table C.1 (continued)

θ	x_0	$\mu(\theta)$	$\sigma^2(\theta)$
0	2	-.464006	1.26067
.2	2	-.269297	1.08614
.4	2	-.279557	1.0287
.6	2	-.291032	1.05463
.8	2	-.277453	1.05132
1	2	-.234538	.951046
1.2	2	-.227617	.872166
1.4	2	-3.35383E-02	.952543
1.6	2	-7.91159E-02	.945992
1.8	2	-.148176	1.03756
2	2	-8.64297E-02	.941652
2.2	2	-5.33609E-02	1.02348
2.4	2	1.49513E-02	.91955
2.6	2	6.37311E-03	.99689
2.8	2	-1.87182E-02	.982744
3	2	-9.40916E-02	1.02233
3.2	2	-4.70771E-02	.958762
3.4	2	9.29452E-02	.885614
3.6	2	-6.78205E-02	1.0096
3.8	2	-3.76485E-02	.945728
4	2	-4.36601E-02	.919874

Table C.2

Mean and Variance of the Pierce Ancillary

θ	x_0	$\mu(\theta)$	$\sigma^2(\theta)$
0	1	.908901	1.74458E-02
.2	1	.889083	2.21687E-02
.4	1	.875867	2.64931E-02
.6	1	.834694	.036621
.8	1	.782491	5.06133E-02
1	1	.729971	.058487
1.2	1	.616717	6.67731E-02
1.4	1	.579689	7.19714E-02
1.6	1	.467329	.065711
1.8	1	.400459	6.06354E-02
2	1	.362156	.06212
2.2	1	.30111	.050483
2.4	1	.243186	3.74997E-02
2.6	1	.199728	2.30171E-02
2.8	1	.160282	1.10205E-02
3	1	.130625	7.33953E-03
3.2	1	.113274	4.49196E-03
3.4	1	9.51982E-02	1.73191E-03
3.6	1	8.34282E-02	4.475E-04
3.8	1	.073889	4.06992E-04
4	1	6.54833E-02	1.49704E-04

REFERENCES

Aitchison, J., and Dunsmore, I.R. (1975), Statistical Prediction Analysis, London: Cambridge University Press.

Barnard, G.A., and Sprott, D.A. (1971), "A Note on Basu's Examples of Anomalous Ancillary Statistics," in Godambe and Sprott (1971), 163-176.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978), Statistics for Experimenters, New York: John Wiley.

Barndorff-Nielsen, O. (1978), Discussion of Efron and Hinkley (1978).

Barndorff-Nielsen, O. (1980), "Conditionality Resolutions," Biometrika, 67, 293-310.

Barndorff-Nielsen, O. (1982), Discussion of Buehler (1982).

Brier, G.W.. (1950), "Verification of Forecasts Expressed in Terms of Probability," Monthly Weather Review, 78, 1-3.

Buehler, R.J. (1959), Some Validity Criteria for Statistical Inferences," Annals of Mathematical Statistics, 30, 845-863.

Buehler, R.J. (1971), "Measuring Information and Uncertainty," in Godambe and Sprott (1971), 330-341.

Buehler, R.J. (1982), "Some Ancillary Statistics and Their Properties," (with discussion and rejoinder), Journal of the American Statistical Association, 77, 581-594.

Cox, D.R. (1958), "Some Problems Connected with Statistical Inference," Annals of Mathematical Statistics, 29, 357-372.

Cox D.R., and Hinkley, D.V. (1974), Theoretical Statistics, London: Chapman and Hall.

Efron, B. (1975), "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," Annals of Statistics, 3, 1189-1242.

Efron, B., and Hinkley, D.V. (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information," Biometrika, 65, 457-487.

References continued

Fisher, R.A. (1925), "Theory of Statistical Estimation," Proceedings of the Cambridge Philosophical Society, 22, 700-725.

Fisher, R.A. (1934), "Two New Properties of Mathematical Likelihood," Proceedings of the Royal Society, Ser. A., 144, 285-307.

Godambe, V.R., and Sprott, D.A. (1971), "Foundations of Statistical Inference." Toronto: Holt, Rinehart and Winston.

Good, I.J. (1952), "Rational Decisions," Journal of the Royal Statistical Society, Ser. B, 14, 107-114.

Hendrickson, A.D., and Buehler, R.J. (1971), "Proper Scores for Probability Forecasters," Annals of Mathematical Statistics, 42, 1916-1921.

Hinkley, D.V. (1977), "Conditional Inference About a Normal Mean with Known Coefficient of Variation," Biometrika, 64, 105-108.

Hora, R.B., and Buehler R.J. (1967), "Fiducial Theory and Invariant Prediction," Annals of Mathematical Statistics, 38, 795-801.

Lehmann, E.L. (1959), Testing Statistical Hypotheses, New York: John Wiley.

Lindley, D.V. (1971), "The Estimation of Many Parameters," in Godambe and Sprott (1971), 435-455.

Pierce, D.A. (1975), Discussion of Efron (1975).

Peers, H.W. (1978), "Second Order Sufficiency and Statistical Invariants," Biometrika 65, 489-496.

Savage, L.J. (1971), "The Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, 66, 783-806.